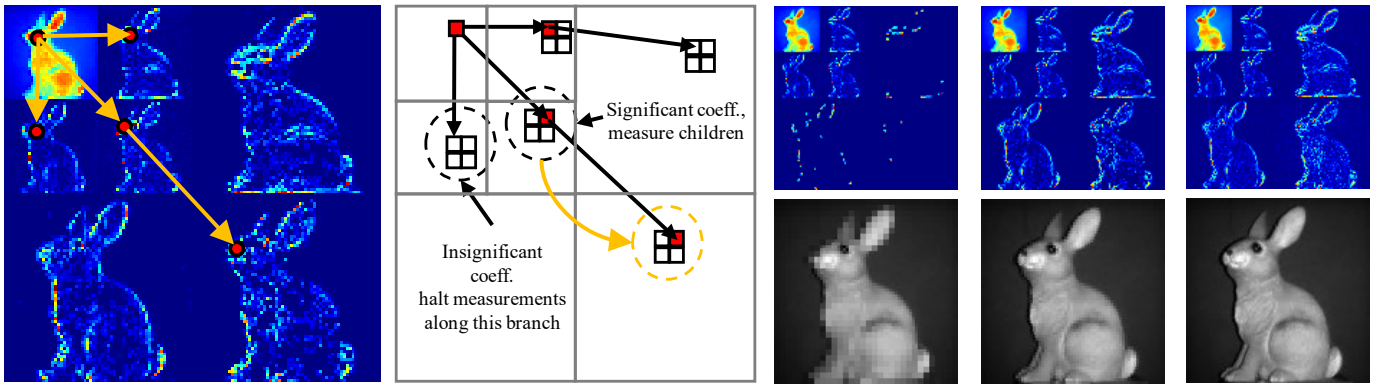
Wavelet Tree Parsing with Freeform Lensing

Vishwanath Saragadam and Aswin C. Sankaranarayanan
Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA

We propose an architecture for adaptive sensing of images by progressively measuring its wavelet coefficients. Our approach, commonly referred to as wavelet tree parsing, adaptively selects the specific wavelet coefficients to be sensed by modeling the children of dominant coefficients to be dominant themselves. A key challenge for practical implementation of this technique is that the wavelet patterns, especially at finer scales, occupy a tiny portion of the field of view and, hence, the resulting measurements have very poor light levels and signal-to-noise ratios (SNR). To address this, we propose a novel imaging architecture that uses a phase-only spatial light modulator as a freeform lens to concentrate a light source and create the wavelet patterns. This ensures that the SNR of measurements remain constant across different spatial scales. Using a lab prototype, we demonstrate successful reconstruction on a wide range of real scenes and show that concentrating illumination enables us to outperform non-adaptive techniques as well as adaptive techniques based on traditional projectors.

***Index Terms*—computational photography, wavelet, adaptive, freeform lens, compressive sensing**



(**a**) Wavelet zero-trees (**b**) Adaptive wavelet-tree sensing (**c**) **Proposed** 12.5% (**d**) 25% (**e**) 50%

Fig. 1: Adaptive wavelet-tree sensing. (a) The 2D wavelet transform of an image is sparse and is accurately modeled as a connected rooted subtrees across spatial scales. (b) This model is used to adaptively sense an image by parsing the children of the dominant wavelet coefficients. We propose an active illumination setup where a phase-only spatial light modulator is used as a freeform lens to concentrate the illumination source onto the projected wavelet patterns. This results in a constant measurement SNR at all spatial scales. (c) - (e) show captured images with our method at various fractions of measurements, illustrating progressive reconstruction.

I. Introduction

THE vast majority of imaging techniques rely on acquiring non-adaptive measurements of the signal of interest. From the classical approach of Nyquist sampling to recent advances in compressive sensing, non-adaptive measurement operators have largely dominated both the theory and implementation of imaging systems. The hallmark of non-adaptive imaging is that the measurement operator is pre-determined. This is in sharp contrast to adaptive imaging where-in each measurement can influence the design of subsequent measurement operators. This paper provides a design for an adaptive imaging system.

Of specific interest to this paper is the so called wavelet tree model, which suggests that the wavelet coefficients of an image exhibit "strong persistence across scales" [1], i.e, the coefficients at a discontinuity tend to persist for a wide range of spatial scales. As a consequence, the wavelet transformation of an image is not just well-approximated as a sparse signal, but also highly structured, i.e., the dominant wavelet coefficients are clustered across multiple scales (see Fig. 1). The persistence of coefficients across scales has numerous applications that range from predictive modeling in image compression [2] to its use as a structured regularizer for denoising [3] and compressive sensing [4].

The predictive power of the wavelet tree model finds its utmost impact when used for adaptively sensing an image [5], where we sequentially measure the wavelet coefficients of the image. At each step, we measure the wavelet coefficients at the spatial location of the coefficient with the largest magnitude, but at a finer scale. This measurement strategy is especially effective, often providing high quality images even when only a small fraction of the coefficients have been measured, with little or no computational overhead for the reconstruction. However, since wavelet atoms are spatially compact, only a small part of the field of view is sensed at finer spatial scales; this dramatically reduces the light incident on the sensor and causes the the signal to noise ratio (SNR) of the measurements to reduce. This vastly limits the practical implementation of the technique especially for sensing at high resolutions.

We propose a novel adaptive imaging strategy that delivers

on the promise of wavelet tree parsing by making two key modifications. First, we pursue an active illumination strategy that projects wavelet atoms on to the scenes to measure the wavelet coefficients. Second, we use a phase-only spatial light modulator (SLM) as a freeform lens [6] to redistribute a light source into the shape of a wavelet atom. We show that this enables measurement of the wavelet coefficients without a drop in the SNR across spatial scales, and results in high quality images, with very few measurements (see Fig. 1).

This paper makes the following contributions.

1) *Wavelet-tree parsing with adaptive illumination.* We propose an active-illumination optical system that is capable of adaptively sensing wavelet coefficients using a phase-only SLM.
2) *Reduced computational footprint.* We show that an adaptive strategy provides progressive reconstructions with negligible computational footprint.
3) *Real experiments.* We demonstrate a hardware prototype and capture a diverse set of scenes to establish the efficacy of our method.

Method	Signal model	Recovery algorithm	Comments
Compressive Sensing	Sparse wavelet transform	ℓ_1 minimization	Slow reconstruction, prone to noise folding [15]
	Wavelet tree	Tree-based CS [31], [1]	Slow reconstruction
DMD projector	Greedy tree pursuit	Inverse wavelet transform	Fast reconstruction, degrading SNR at higher spatial scales
Phase SLM projector	**Greedy tree pursuit**	**Inverse wavelet transform**	**Fast reconstruction, constant SNR for all measurements**

TABLE I: Comparison of sensing strategies. CS relies on random projections of the image that are not adaptive to scene. Further, using a prior results often leads to slow recovery time, amplified with the problem of noise folding [15]. In contrast, wavelet-tree parsing directly captures the dominant wavelet coefficients using a greedy but adaptive approach. However, when the scene is illuminated with a DMD-based projector, the SNR degrades for high frequency features such as edges. Our proposed method projects patterns using a phase-only SLM that concentrates all the light into the spatially compact wavelet pattern. This results in fewer measurements, negligible reconstruction time, and a constant SNR independent of scale.

II. Prior Work

The key ideas of this paper revolve around imaging from a small collection of high SNR measurements. This touches upon ideas in multiplexed imaging and compressive sensing; we briefly discuss them here.

A. Multiplexed imaging

One of the seminal works in computational imaging is that the use of multiplexing codes such as Hadamard matrix [7], [8] increases SNR of reconstructions [9], [10], [11], specifically when the noise is independent of the signal. This has found applications in spectrometry [12], hyperspectral imaging [13], and light transport estimation [10]. While Hadamard multiplexing can lead to increased reconstruction quality, it does not lead to any reduction in the number of measurements that need to be acquired.

B. Compressive sensing

Compressive sensing (CS) aims to recover a signal from a set of multiplexed linear measurements that are fewer than its dimensionality [14]. Given an image $\mathbf{x}$, CS seeks to recover signal from several random projections:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is the measurement noise and Φ is the measurement matrix. Since the measurements are fewer than signal dimension, (1) is an underdetermined set of linear equations. To get a unique solution to this underdetermined problem, we regularize the inverse process using signal priors to obtain an optimization problem of the form

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|^2 + \lambda\Gamma(\mathbf{x}), \tag{2}$$

where $\Gamma(\mathbf{x})$ is a prior on the signal class of interest, typically sparsity of the signal under some transformation. The success of CS relies on the availability of a tractable and concise prior for the signal. One such example is the 2D discrete wavelet transform (DWT) of images.

The DWT is an orthonormal transformation that decomposes an image across multiple spatial scales (Fig 1 (a)). The coarsest scale is called the approximation coefficients and represents smooth regions in the image and the finer scales called detail coefficients capture transitions such as edges at different spatial scales. Since images are well approximated as being smooth except for a finite number of discontinuities (edges), the DWT often provides a sparsifying transformation.

a) Single pixel camera: CS of images with real hardware was initiated with the seminal work of single pixel camera (SPC) by Duarte et al. [16]. Instead of using a 2D sensor, images are captured using a single photodetector along with a spatial light modulator (SLM). The SLM is programmable and enables measurement of random linear projections of the scene being imaged, which are then used to reconstruct the image by solving the optimization function in (2). This results in a low-cost optical setup that required only a single measurement unit, and captures far fewer measurements than the image dimension.

b) Projector-based SPC: When the imaging hardware has illumination as an integral component — for example, in microscopy — a projector may be used instead of a 2D spatial light modulator [17]. The same compressive measurements of image of the scene can be obtained by projecting the desired patterns on the scene and measuring the scene light level with a photodetector. Our method is an instance of a projector-based SPC, but with the key difference that we use a phase-only SLM to generate the wavelet patterns, which efficiently redistributes instead of spatially attenuating light.

c) Drawbacks of CS: While CS is a promising technique for high-dimensional images, practical implementations suffer from some setbacks. CS suffers from the problem of noise

folding. Treichler et al. [15] show that for every doubling in compression of measurements, the reconstruction SNR falls by 3 dB. Hence, CS does not provide significant benefits at higher compression ratios. Further, recovery of signal requires a computationally expensive and iterative optimization algorithm. Finally, the measurement strategy of CS is non-adaptive and not tailored to the specific instance of signal being measured. However, there are significant benefits that can be achieved if sampling strategy is tightly coupled to the specific signal being sensed.

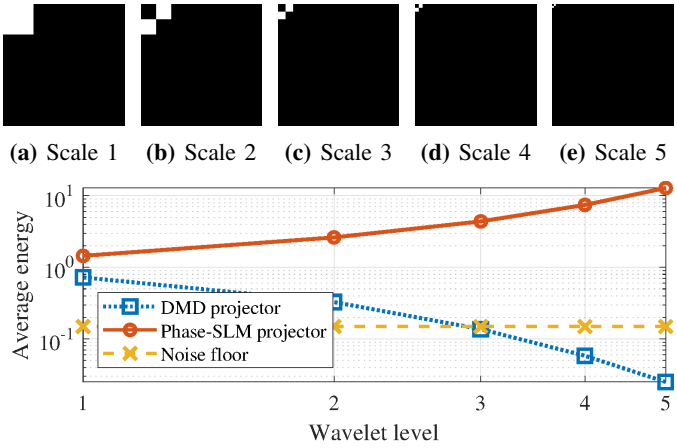
C. Adaptive sensing

In contrast to CS, adaptive sensing relies on measurements tailored to the specific instance of a signal. At the heart of adaptive sensing is a *predictive signal model* that enables prediction of the next measurement based on some or all of the previous measurements. This admits a sensing strategy that maximizes SNR of measurements with fewest measurements, as only the significant components of the signal are sensed. Further, such a sensing scheme often has minimal reconstruction time, as the signal (or its orthonormal transformation) is sensed directly. There is a rich body of literature that relies on adaptive sensing to achieve higher accuracy [18], [19], or to capture fewer measurements [20]. In the field of visual signals, adaptive sensing has shown promising results for estimating light transport matrices [21], hyperspectral images [22], and images [5]. Our paper focuses on adaptively sensing of images using an SPC architecture.

Wavelet trees of images: A key example of a predictive model, and central to our proposed method, is the rooted tree structure of wavelet coefficients of images. Specifically, the wavelet tree model organizes the wavelet coefficients on a tree with each node corresponding to a coefficient and each level of the tree corresponding to the wavelet transform scale. Under such a configuration, the dominant coefficients lie on a rooted subtree, i.e., if a parent coefficient is zero, then all the children coefficients are zero with high probability. The intuition is that discontinuities in the form of edges persist across scales and hence manifest as dominant coefficients. The correlations between wavelet coefficients has been extensively studied before [23], [24], [25] and has been used for significant improvements in image compression [2], [26], rendering [27], and even compressive imaging [4], [28], [29]. In this context, this paper proposes an imaging hardware that relies on wavelet tree of images to capture image of a scene with very few measurements at high SNR.

III. Adaptive Wavelet Tree Parsing

Closely related to the ideas in this paper is the work by Deutsch et al. [5] who also proposed an adaptive imaging scheme using wavelet-tree parsing. The proposed strategy is initialized by sensing all approximation coefficients of the image of the scene using an SPC architecture. Then, the detail coefficients are sampled by traversing the wavelet-tree, where only children of significant parent coefficients are measured (Fig. 1 (b)). When images are well modeled as a wavelet tree, this adaptive sensing scheme outperforms reconstructions from non-adaptive compressive sensing. Further, since the wavelet coefficients are sampled directly, we can recover the sensed image by applying an inverse DWT. The results in Deutsch et al. are, however, all based on simulated data; as we see next, a straightforward implementation of this idea results in measurements that have poor SNR at the finer spatial scales.



(a) Scale 1 **(b)** Scale 2 **(c)** Scale 3 **(d)** Scale 4 **(e)** Scale 5

(f) Energy vs. spatial scale

Fig. 2: Fall-off of energy in wavelet bands. The plot shows average energy of wavelet coefficients simulated on the "cameraman" image. Wavelet patterns at finer spatial scales multiplex a smaller field of view which results in reduced SNR of measurements. This poses a problem when using adaptive sensing, as the measurement noise dominates at high spatial scales. Instead of using a DMD projector to generate patterns, we use a phase-only SLM as a freeform lens to concentrate light over the wavelet basis's spatial support, thus maintaining measurement energy above noise floor. Then adaptive sensing of images results in higher SNR than CS.

A. Failure of projector-based wavelet-tree sensing

A practical implementation of adaptive imager proposed in [5] is impeded by noise in measurements. Since wavelet bases are spatially compact, a smaller field of view of the scene is multiplexed resulting in reduced signal energy at finer scales. This leads to a decreasing SNR of the measurements with increasing spatial scales. To be specific, the average SNR drops by a factor of four with a doubling of the scale parameter. This can be seen in the plot of average energy as a function of spatial scale in Fig. 2 simulated on the "cameraman" image. Given a noise floor, the measurements at finer scales dramatically reduced SNR, which results in reduced reconstruction quality at high spatial resolutions. Alternatively, acquiring measurements with a constant SNR across scales would require an exponential increase in the exposure time. Our core contribution is an imaging hardware that concentrates all light within the spatially compact region to maintain a constant measurement SNR.

B. Freeform lensing for concentrating illumination

The effect of reduced signal levels with increasing spatial scales can be overcome if we use a longer exposure duration to collect more photons on the sensor. However, such a strategy

(a) Phase SLM, 6.25% 22.8dB **(b)** 12.5%, 25.3dB **(c)** 25%, 26.6dB **(d)** 50%, 27.2dB **(e)** 100%, 27.9dB

(a) DMD projector, 6.25% 22.4dB **(b)** 12.5%, 23.4dB **(c)** 25%, 22.6dB **(d)** 50%, 20.4dB **(e)** 100%, 17.8dB

(a) CS, 6.25% 17.2dB **(b)** 12.5%, 18.8dB **(c)** 25%, 20.9dB **(d)** 50%, 21.8dB **(e)** 100%, 22.5dB

Fig. 3: Imaging with various strategies. We simulated imaging on the house image (256×256) with phase SLM projector (first row), DMD projector (second row), and CS (third row). CS was performed with permuted Hadamard measurements and recovered with sparsity in the wavelet basis. We modeled noise as photon and readout noise. At low resolution, the advantages of concentrating light are minimal, since signal energy is above noise floor. However, adaptive illumination outperforms both DMD projector and CS at higher resolution.

requires exponentially longer time for capturing higher resolution wavelet coefficients. Instead, we use a phase-only SLM to project wavelet atoms on the scene. A phase-only SLM generates patterns by redistributing light intensity within the spatial support of a pattern and can be seen as a programmable freeform lens [6]. Hence the total energy in the pattern stays the same independent of size or shape of the desired pattern. This results in a constant SNR of measurements at all scales, with no increase in measurement time. The plot in Fig. 2 (red solid line) shows the energy of wavelet coefficients when using an adaptive illumination strategy illustrating that the energy stays well above the noise floor at all spatial scales. Coupled with wavelet-tree parsing idea, we demonstrate the first adaptive imager with high quality images captured with only a few measurements.

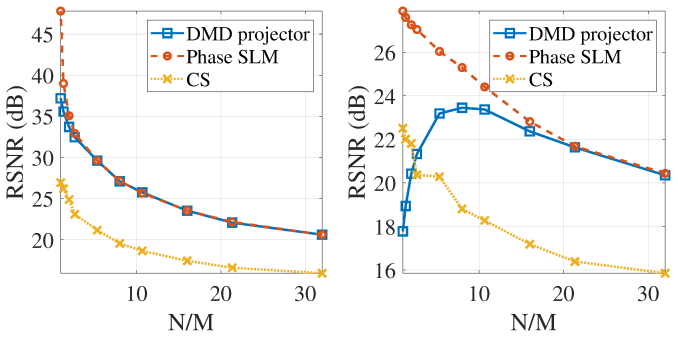
C. Simulations

We compare our method against compressive sensing with passive wavelet-tree parsing as well as compressive sensing using permuted Hadamard entries. We model each measurement as $y_k = \mathcal{P}(\tau \boldsymbol{\phi}_k^\top \mathbf{x}) + n$, where τ models intensity of measurements, $\mathcal{P}$ captures Poisson noise, $n \sim \mathcal{N}(0, \sigma^2)$ models readout noise, and $\boldsymbol{\phi}_k$ is the pattern projected on the scene. The values of $\mathbf{x}$ were normalized between $0, 1$ and used a readout noise of 50 dB to emulate machine vision cameras. Henceforth, we refer to the adaptive wavelet-tree sensing using a projector as "DMD projector" sensing strategy,[1] and the proposed method as "Phase SLM" projector sensing strategy. For simulations, as well as in our real experiments, we used a separable Haar wavelet transform. The results remain the same for other wavelet bases. Sensing using phase SLM as a projector was emulated by scaling value of the projected pattern. Hence, when simulating projection of a pattern at scale J, the amplitude is 4^J instead of 1. Finally, CS recovery was done using ℓ_1 minimization of wavelet transform.

Figure 3 shows simulations of imaging using a DMD and phase SLM for varying measurement rates. Using a phase SLM outperforms a DMD projector strategy, as well as compressive sensing for any number of measurements which is in agreement with our hypothesis. Moreover, the SNR of DMD sensing degrades with increasing measurements, as the higher scale coefficients are very noisy and add to inaccuracies.

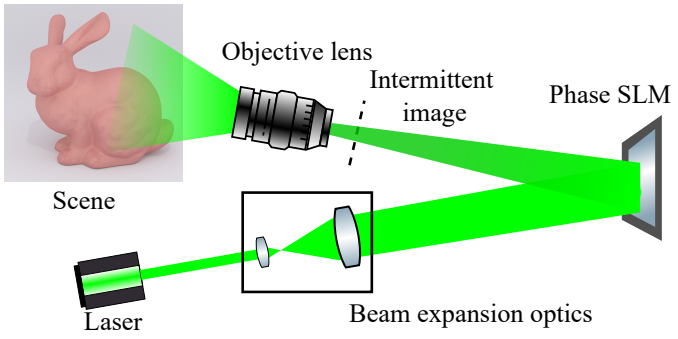
It is worth noting that the gains of concentrating light are

[1]The usage of DMD projector is only for ease of exposition. The main idea is that a traditional projector that uses DMD or Liquid Crystal on Silicon (LCoS) operates by attenuating amplitude, while we only modify the phase.

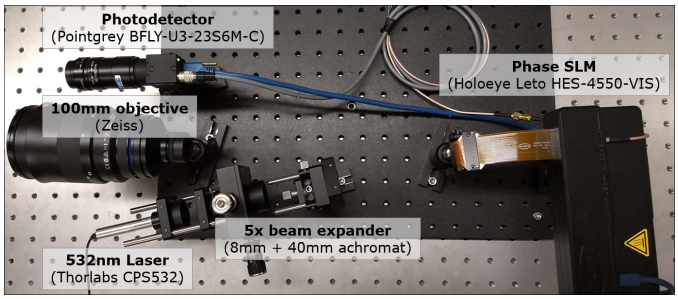


(a) Bright projector **(b)** Dim projector

Fig. 4: Comparison of various imaging techniques. The plots above show accuracy vs compression for the cameraman image when imaged with a (a) bright and (b) dim projector. We observe using a phase SLM provides significant gains when operating with sources that are not bright.



(a) Optical schematic



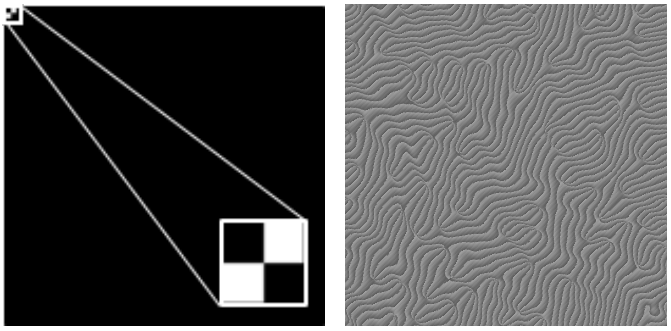
(b) Our prototype

Fig. 5: Schematic and the prototype. The proposed optical setup consists of a monochromatic light source along with a phase-only SLM, which only redistributes light without attenuating amplitude. We used a phase-only SLM from Holoeye with a spatial resolution of 1920×1080. The SLM is capable of introducing a phase shift of up to 2π radians. We used a machine vision camera from Pointgrey instead as a photodetector.

tangible only when measurements are made with light sources that are not too bright, as well as the images are captured at high resolution. Figure 4 shows quantitative plot of various imaging techniques with projector light sources of different intensities. When the scene is illuminated by a bright projector ($\tau = 10{,}000$), concentrating light does not offer any advantage – this is a direct consequence of a low noise floor. However, with low light projector ($\tau = 100$), phase SLM strategy outperforms DMD projector strategy. As expected, the gains are negligible at high compression as the measured coefficients belong to small spatial scales that have large spatial supports.

IV. Implementation Details

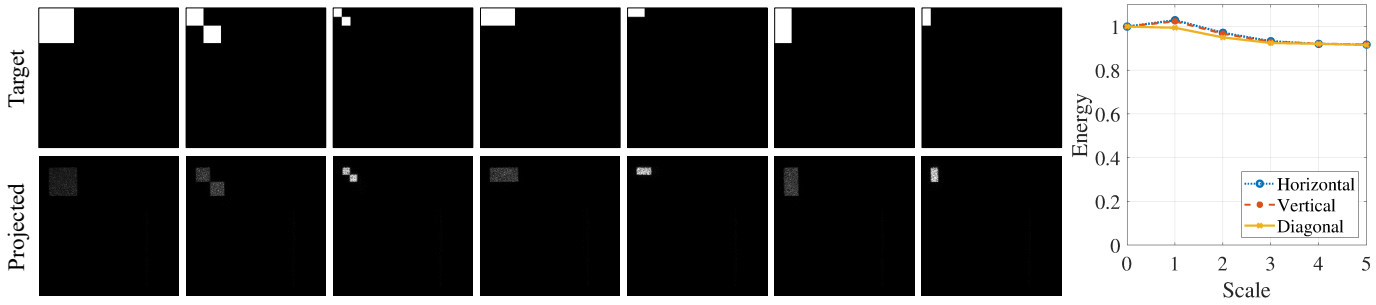
Figure 5(a) shows a schematic of the proposed imaging hardware consisting of a collimated light source, a phase-only SLM, an objective lens, and a camera that we use as a photo-diode. The collimated light is transformed by the phase SLM to display a pattern on the intermittent plane using Fraunhoffer diffraction [30]. Hence, given a pattern, $p(x, y)$, we display its Fourier transform $g(u, v) = \mathcal{F}(p(x, y))$ on the SLM. However, since the SLM only accepts phase values, we require $|g(u, v)| = 1$. We solve for the phase pattern using the Iterative Fourier Transform (IFT) technique [31] (see Fig. 6). The algorithm iteratively satisfies the spatial domain constraint, $|\mathcal{F}(g(u, v))|^2 = p(x, y)$ and frequency domain constraint, $\mathcal{F}(p(x, y)) = g(u, v), \quad |g(u, v)| = 1$. We observe that the optimization process converges to a satisfactory result in 40 iterations and takes approximately 10s for a pattern of size 1024×1024. Computations were done on a workstation with Intel Xeon processor with six physical cores, 64GB RAM and 2TB hard disk.



(a) Pattern **(b)** Phase mask

Fig. 6: Design of phase pattern. We use Iterative Fourier Transform (IFT) to design phase masks. (a) shows positive part of a diagonal wavelet basis and (b) shows the obtained phase mask. Since wavelet bases of a given scale are simple translations of each other, we generate a single pattern and add an appropriate phase ramp to obtain its translated versions.

a) Efficient generation of patterns: While the patterns can all be generated individually and stored before hand, the number of patterns scales quadratically with image resolution. However, since the wavelet atoms at a given scale are translations of each other, we can translate the phase mask by adding a 2D phase ramp whose slope is determined by the amount of translation. Therefore, we only need to generate three positive, and three negative patterns per scale for each of the horizontal, vertical and diagonal wavelet coefficients, and one more for approximation coefficients at the coarsest scale. For imaging at J distinct scales, this requires optimization of a total of $6J + 1$ patterns. Figure 6 shows the desired pattern, generated phase mask and the projected pattern for some representative examples. Figure 7 shows example of projector patterns as well as the measured energy in each of the generated pattern. As expected, the energy of wavelet patterns formed on the



(a) Target patterns and projected patterns **(b)** Energy of each pattern

Fig. 7: Energy vs. scale. A phase SLM forms patterns by redistributing light, and hence the energy stays constant, independent of shape or scale of pattern. (a) shows target and projected patterns for some Haar wavelet patterns. Note how the intensity in smaller wavelet patterns is higher. (b) shows energy of each pattern as a function of spatial scale, which is almost constant.

scene is approximately constant.

b) Adaptation strategy: Deutsch et al. [5] proposed an adaptive strategy where children of coefficients above a fixed threshold are measured starting from coarse scale to fine scale, and parsing each rooted sub-tree. The same strategy is also used in JPEG2000 for compressing images using wavelet trees [2]. However, such strategies are not be suitable for a given measurement budget, as the mapping from a given threshold to number of measurements is scene-dependent.

Instead, we use a greedy approach of scanning, where the next set of measurements are the three children of the measured coefficient with maximum magnitude. Specifically, let I_{DWT} be the currently measured set of wavelet coefficients, and let $C((x,y))$ be the children indices of DWT index (x,y). The proposed greedy approach measures $C(\arg\max_{x,y} |I_{\text{DWT}}(x,y)|)$. The algorithm then proceeds by finding the next maximum, and so on. The algorithm is encapsulated in Algorithm 1. Such a greedy strategy is motivated by the idea that coefficients of edges have large magnitude, and hence are sensed earlier than the smooth regions. The proposed approach is illustrated with a simple diagonal image in Fig. 8. When measurements are only a few, the algorithm prioritizes low resolution coefficients, with some emphasis on edges. With increasing number of measurements, priority is given to the diagonal edge, visible as non-zero wavelet coefficients in higher scales.

Algorithm 1 Greedy DWT measurement. PROJECT() displays a pattern on the phase SLM. MEASURE() captures a measurement from photo detector. CHILDREN() returns children indices of a given DWT index. DWT_LOC() computes shift and scale from DWT index.

Require: Number of measurements N, Initial resolution J
  Initialization
  $I_{\text{DWT}} \leftarrow \text{ZEROS}(H, W)$ ▷ Initialize DWT image
  $k \leftarrow 0$ ▷ Current number of measurements

  Step 1: Measure approximation coefficients
  for $x = 1 \cdots J$ **do**
    for $y = 1 \cdots J$ **do**
      PROJECT($\phi(x,y)$) ▷ Project pattern
      $I_{\text{DWT}}(I_{\text{meas}}) \leftarrow$ MEASURE() ▷ Photo detector
      $k \leftarrow k + 1$
    end for
  end for

  Step 2: Measure analysis coefficients
  while $k \leq N$ **do**
    $x_{\max}, y_{\max} \leftarrow \arg_{x,y} |I_{\text{DWT}}(x,y)|$ ▷ Get largest meas.
    $\mathcal{I}_{\text{meas}} \leftarrow \text{CHILDREN}(x_{\max}, y_{\max})$ ▷ Get child. ind.
    for $I_{\text{meas}} \in \mathcal{I}_{\text{meas}}$ **do**
      $x_0, y_0, j \leftarrow \text{DWT\_LOC}(I_{\text{meas}})$ ▷ Get DWT params
      PROJECT($\psi_j(x - x_0, y - y_0)$) ▷ Project pattern
      $I_{DWT}(x,y) \leftarrow$ MEASURE() ▷ Photo detector
      $k \leftarrow k + 1$
    end for
  end while

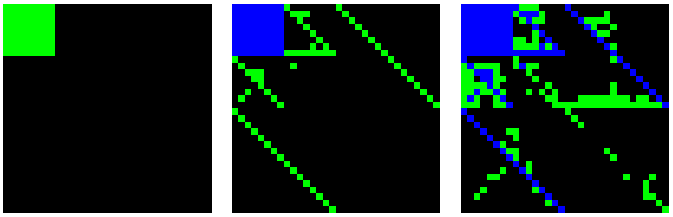
  return I_{DWT} ▷ Return captured DWT

V. Real Experiments

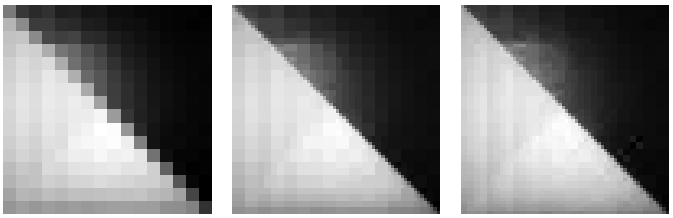
We evaluate the proposed method by capturing images of a diverse set of scenes. We show results with raster scanning and wavelet-tree parsing at a spatial resolution of 128×128, and some representative examples for wavelet-tree parsing at 256×256 spatial resolution.

a) Progressive reconstruction: Figure 1 compares wavelet transform of an image captured by raster scanning the scene, and the wavelet coefficients captured by optically scanning the wavelet tree. We make two observations here. First, our greedy strategy pursues discerning features such as edges, as is evident from transition from image (c) to (e). Second, the quality of reconstruction gets progressively better with increasing number of measurements and matches the wavelet transform computed on raster-scanned image.

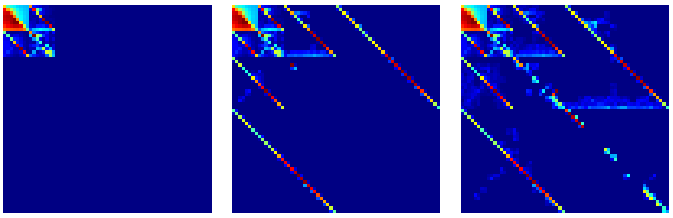
b) Accuracy comparison: Figure 9 shows two example scenes. The snowman scene consists of a soft toy with some spatial texture and some strong edges. The features of the toy, such as mouth, nose and eyes are well resolved even with 25% measurements. The toys scene on the other hand has more spatial complexity, and hence requires up to 50% measurements for an accurate reconstruction. The images are of good quality even at high compression due to concentration of sampling around edges of the image.



(a) Newly added samples (green)



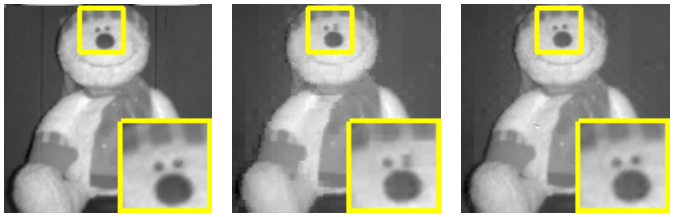
(b) Recovered image



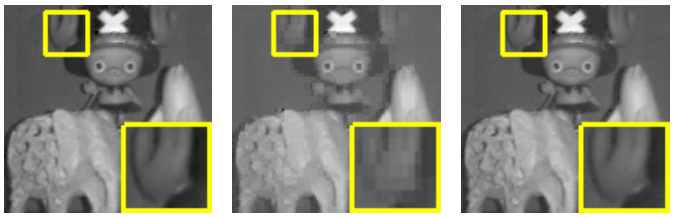
(c) 256 meas. **(d)** 512 meas. **(e)** 1024 meas.

(f) Wavelet coefficients

Fig. 8: Illustration of sampling strategy. We incorporate a greedy sampling strategy where children of large magnitude coefficients are measured first. Top row shows newly added samples in green, previously captured samples in blue and unsampled coefficients in black. Middle row shows reconstructed image and bottom row shows corresponding wavelet transform. At low sampling rate, the coarse scale coefficients are given priority. As sampling rate increases, the edge coefficients, visible as diagonal streaks in higher scales are prioritized. This enables a progressive reconstruction of images, while giving high importance to discerning features such as strong edges.



(a) Snowman **(b)** 25% wavelet, $16.2dB$ **(c)** 100% wavelet, $16.5dB$



(d) Toys **(e)** 25% wavelet, $20.2dB$ **(f)** 100% wavelet, $21.4dB$

Fig. 9: Performance of proposed method for varying compression. The proposed method is capable of time-budgeted measurements. The quality progressively improves with increasing number of measurements, similar in spirit to progressive decompression of JPEG images [2]. The results shown here are on 128×128 images with varying number of measurements. The accuracy is close to maximum even at low sampling rates.

c) Comparison with CS: Figure 10 shows comparison between proposed method and CS-based image reconstruction for varying number of measurements. The images were captured at 64×64 spatial resolution. CS measurements were captured with subsampled and permuted Hadamard matrix, and image was recovered using sparsity in wavelet basis. Imaging using adaptive sensing looks visually superior to CS results at all compression rates. Furthermore, while CS recovery took over a minute for each image, adaptive sensing requires practically no time to recover the image.

d) Comparison with DMD projector: Figure 11 shows results of adaptive wavelet tree sensing using a DMD projector as well as our proposed hardware. We observe that light concentration via freeform lensing does provide reconstructions at higher quality.

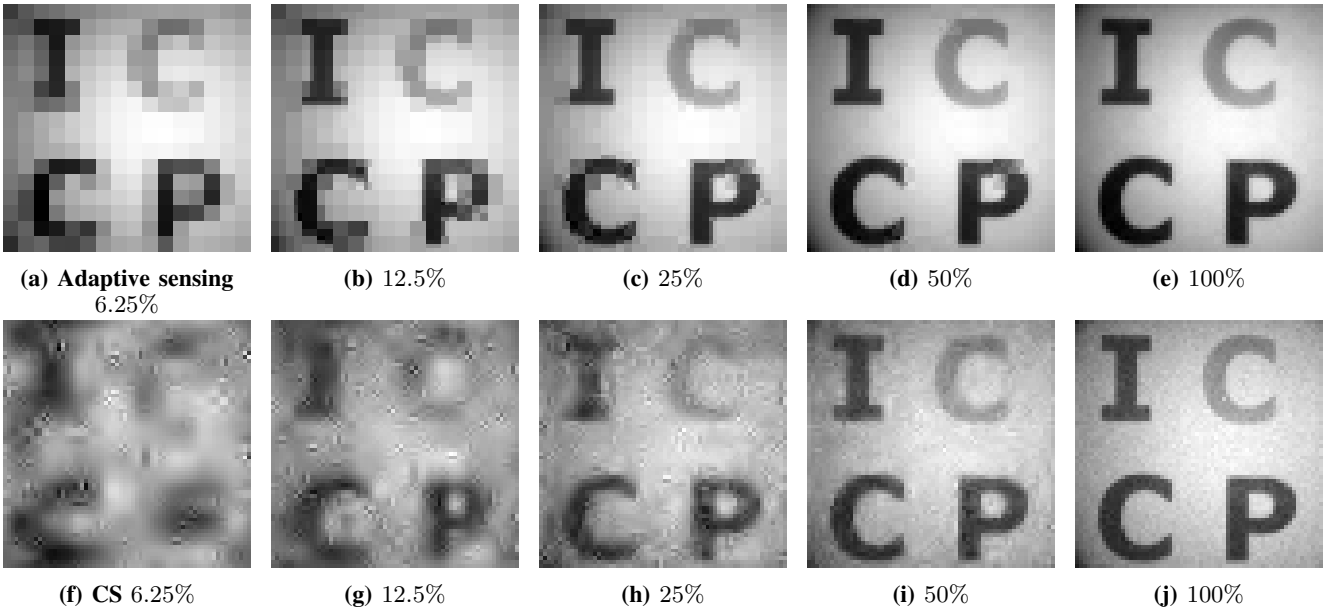
e) Advantages at high resolution: Since images are sparser at higher resolution, adaptive sensing requires a smaller fraction of measurements to recover high quality images. Figure 12 shows images of horse bust at 128×128 resolution, with 100% wavelet samples captured, as well as a 256×256 image with only 25% samples captured. In particular, the details around horse's ear, as well as its mane is well resolved in 256×256 image.

f) Wavelet foveation: Foveation of a scene can be achieved by selecting a region of interest from the low-resolution approximation image, and then giving high weights to this region. We show an example of foveated sampling in Fig. 13 where the region of interest is manually selected to be over the marble bust's face. Foveation is done by first measuring all approximation coefficients first, and then using the proposed greedy wavelet sampling strategy within the region of interest. Fine details like the curl of hair are highly resolved within the region. The compression in measurements was $5\times$ for fully sampling within the region and $10\times$ for greedy-adaptive sampling, clearly underlying the advantages of our method.

g) Sampling with other wavelet bases: While we showed all our results with 2D separable Haar wavelet transform, the results hold for any other wavelet bases. To illustrate this, we generated Daubechies-2 wavelet bases and performed wavelet-tree sensing on a representative scene. Figure 14 shows comparison of the proposed method with Haar wavelet basis and Daubechies-2. As is to be expected, the results are identical at 100% sampling (b, d). However, sampling at lower sampling rate (a, c) shows blocky artifacts with Haar wavelets and smoother artifacts with Daubechies-2. Depending on the specific scenes being captured, any of the wavelet bases may be chosen for adaptive imaging.

VI. Discussions and Conclusion

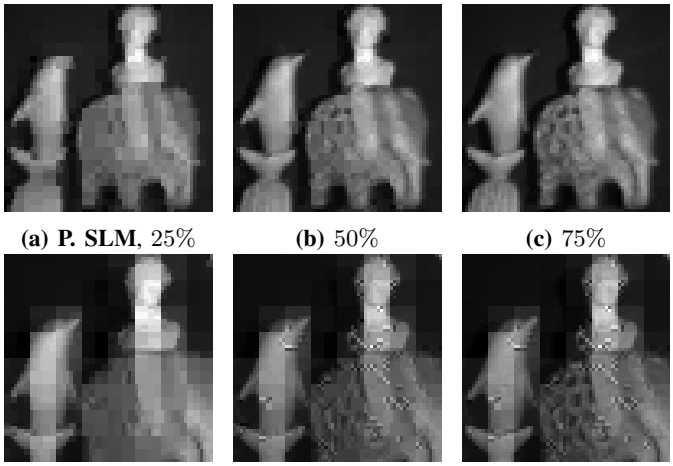
We proposed and evaluated a new imaging hardware that is capable of measuring images at high resolution at high



(a) Adaptive sensing 6.25% **(b)** 12.5% **(c)** 25% **(d)** 50% **(e)** 100%

(f) CS 6.25% **(g)** 12.5% **(h)** 25% **(i)** 50% **(j)** 100%

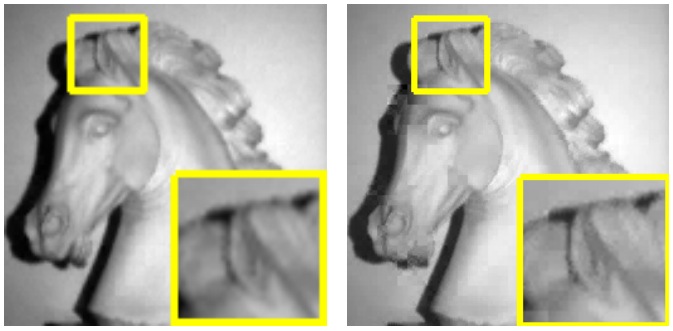
Fig. 10: Progressive reconstruction with CS and adaptive sensing. A key advantage of adaptive sensing is that the results get progressively better with minimal computation time. We show an example of a 64×64 scene made of letters "ICCP" with reconstruction using adaptively-sensed measurements, and CS with permuted subsampled Hadamard multiplexing and recovery using ℓ_1 sparsity in wavelet domain. The visual quality of adaptive sensing is superior to compressive sensing at all measurement rates. Further, while adaptive sensing takes negligible amount of time to reconstruct images, we need to solve a complex optimization problem for CS everytime we get new measurements.



(a) P. SLM, 25% **(b)** 50% **(c)** 75%

(d) DMD p., 25% **(e)** 50% **(f)** 75%

Fig. 11: DMD projector vs phase SLM. We captured a 64×64 image with a laser projector as well as our prototype to compare the advantages of concentrating light. While the images look similar at low sampling rates, adaptive sensing with DMD adds high frequency artifacts at higher sampling rates, a direct consequence of reduced signal energy.

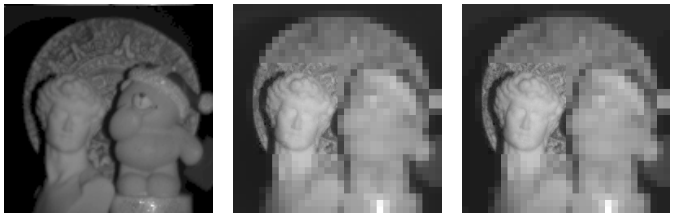


(a) 128×128, 100% **(b)** 256×256, 25%

Fig. 12: Advantages at higher resolution. Images are sparser as resolution increases, which works in favor of adaptive wavelet-tree parsing. We show an example scene comprising of a horse bust sampled at 128×128 as well as at 256×256 with same number of measurements. Despite a $2\times$ increase in resolution, we achieve finer spatial details without an increase in measurements.

compression rates. This is achieved by projecting wavelet patterns using a phase-only SLM which concentrates light into the spatially compact wavelet pattern, thereby overcoming reducing SNR with higher resolution. We captured real results over a diverse set of scenes, showing the superiority of our method in terms of fewer measurements and higher quality of reconstruction. Our method for the first time shows decoupling of image resolution and number of samples, without long reconstruction times or compromise in image quality.

a) Limitations: The proposed method has limitations that arise from three sources — the use of active illumination, failure of wavelet tree models and finally, the specific prototype used. First, the use of an active illumination limits the range of scenes on which our imager can be used; scenes that contain strong global effects and specular reflections violate the imaging model used in the paper and hence, lead to artifacts in reconstructions (see Fig. 15). Second, images with high frequency features have a dense wavelet transform that do not necessarily follow the wavelet tree model, requiring a full parsing of the images' DWT. Figure 16 shows a scene captured by our setup at 128×128 spatial resolution with varying number of measurements and the corresponding

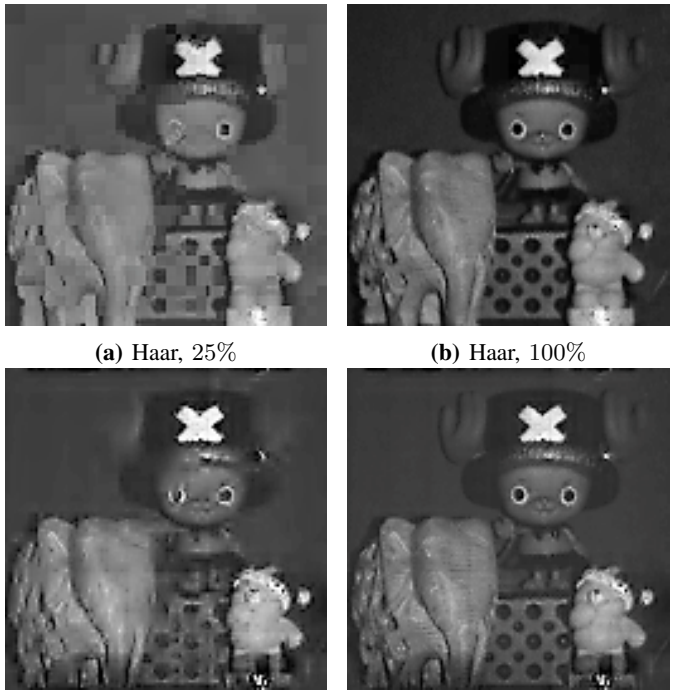


(a) 128×128 raster scan

(b) 256×256 foveated, 10% measurements

(c) 256×256 foveated, 19% measurements

Fig. 13: Optical wavelet foveation. Our setup is capable of selectively sampling a region of interest; We show an example here with selectively high resolution sampling over the marble bust. Even with 10% measurements, our method is capable of capturing fine details in the scene.



(a) Haar, 25%

(b) Haar, 100%

(c) Daubechies-2, 25%

(d) Daubechies-2, 100%

Fig. 14: Sampling with Daubechies wavelets. Though Haar wavelets are a popular choice, our setup is capable of capturing any wavelet decomposition. Here we show a comparison with Daubechies wavelets with two filter taps.

wavelet transform. The key observation here is that the central high frequency part is not sampled till much later. The sector star is a good example of violation of the wavelet-tree model – the coefficients at lower spatial scales are zero, but are significantly high at higher spatial resolution. Third, our current prototype is bottlenecked by speed of phase-only SLM that runs at 10fps. This limits real-time imaging based on wavelet sensing. One potential way to speed up acquisition is use galvos and focus tunable lenses to translate and scale the projected pattern, thereby avoiding the bottlenecks presented by the low operating speed of the SLM.

(a) Frog imaged by cellphone camera

(b) Wavelet, 25%

(c) Wavelet, 100%

Fig. 15: Failure due to non-Lambertian BRDF. When the scene to be imaged is highly specular, active illumination based methods fail to work well. This can be seen from the blocky artifacts on the lips of the ceramic frog.

VII. Acknowledgements

The authors acknowledge support from the National Science Foundation under the grants CCF-1652569, IIS-1618823, and the National Geospatial-Intelligence Agency's Academic Research Program (Award No. HM0476-17-1-2000).

References

[1] R. G. Baraniuk, "Optimal tree approximation with wavelets," in *Wavelet Applications in Signal and Image Processing VII*, 1999.

[2] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[3] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[4] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[5] S. Deutsch, A. Averbush, and S. Dekel, "Adaptive compressed image sensing based on wavelet modeling and direct sampling," in *SAMPTA*, 2009, pp. General–session.

[6] G. Damberg, J. Gregson, and W. Heidrich, "High brightness HDR projection using dynamic freeform lensing," *ACM Trans. Graphics*, vol. 35, no. 3, p. 24, 2016.

[7] M. Harwit, *Hadamard Transform Optics*. Elsevier, 2012.

[8] M. Harwit and N. J. Sloane, *Hadamard Transform Optics*, 1979.

[9] O. Cossairt, M. Gupta, and S. K. Nayar, "When does computational imaging improve performance?" *IEEE Trans. Image Processing*, vol. 22, no. 2, pp. 447–458, 2013.

[10] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, "A theory of multiplexed illumination," in *Intl. Conf. Computer Vision*, 2003.

[11] Y. Y. Schechner, S. K. Nayar, and P. Belhumeur, "Multiplexing for optimal lighting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1339–1354, 2007.

[12] J. A. Decker, "Experimental realization of the multiplex advantage with a Hadamard-transform spectrometer," *Applied Optics*, vol. 10, no. 3, pp. 510–514, 1971.

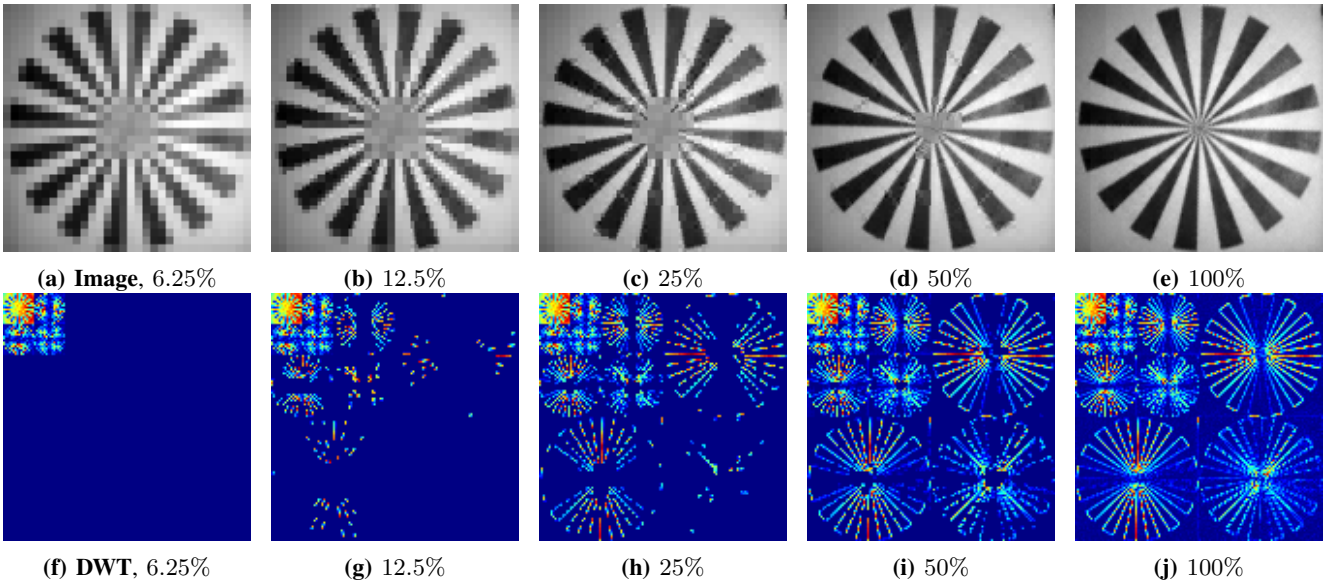
[13] M. Gehm, M. Kim, C. Fernandez, and D. Brady, "High-throughput, multiplexed pushbroom hyperspectral microscopy," *Optics Express*, vol. 16, no. 15, pp. 11 032–11 043, 2008.

[14] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[15] J. Treichler, M. Davenport, and R. Baraniuk, "Application of compressive sensing to the design of wideband signal acquisition receivers," *US/Australia Joint Work. Defense Appl. Signal Processing*, vol. 5, 2009.

[16] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, p. 83, 2008.

[17] F. Magalhães, F. M. Araújo, M. V. Correia, M. Abolbashari, and F. Farahi, "Active illumination single-pixel camera based on compressive sensing," *Applied Optics*, vol. 50, no. 4, pp. 405–414, 2011.



(a) Image, 6.25% **(b)** 12.5% **(c)** 25% **(d)** 50% **(e)** 100%

(f) DWT, 6.25% **(g)** 12.5% **(h)** 25% **(i)** 50% **(j)** 100%

Fig. 16: Failure case of wavelet tree parsing. We show an example of progressive reconstruction for the sector star image which is used for evaluating modulation transfer function (MTF). The central high frequency zone does not get sampled even after 50% measurements. This is one of the failure cases of greedy tree algorithm – the sector star image is not accurately represented by the wavelet tree structure as the wavelet coeffficients of the center of sector have smaller magnitude at coarser scale but higher magnitude at finer scales.

[18] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *IEEE Trans. Info. Theory*, vol. 57, no. 9, pp. 6222–6235, 2011.

[19] J. Haupt, R. Nowak, and R. Castro, "Adaptive sensing for sparse signal recovery," in *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, 2009, pp. 702–707.

[20] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," *IEEE Trans. Info. Theory*, vol. 60, no. 7, pp. 4001–4012, 2014.

[21] M. O'Toole and K. N. Kutulakos, "Optical computing for fast light transport analysis," *ACM Trans. Graphics*, vol. 29, no. 6, p. 164, 2010.

[22] V. Saragadam and A. C. Sankaranarayanan, "KRISM–Krylov subspace-based optical computing of hyperspectral images," *arXiv Preprint arXiv:1801.09343*, 2018.

[23] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[24] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Analysis and Machine Intelligence*, no. 7, pp. 710–732, 1992.

[25] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE trans. info. theory*, vol. 38, no. 2, pp. 617–643, 1992.

[26] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM J. Mathematical Analysis*, vol. 29, no. 2, pp. 511–546, 1998.

[27] R. S. Overbeck, C. Donner, and R. Ramamoorthi, "Adaptive wavelet rendering." *ACM Trans. Graphics*, vol. 28, no. 5, pp. 140–1, 2009.

[28] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 57, no. 9, pp. 3488–3497, 2009.

[29] C. Hegde, P. Indyk, and L. Schmidt, "A fast approximation algorithm for tree-sparse recovery," in *IEEE Intl. Symposium on Info. Theory*, 2014.

[30] J. W. Goodman, *Introduction to Fourier Optics*. Roberts and Company Publishers, 2005.

[31] J. Liu and M. Taghizadeh, "Iterative algorithm for the design of diffractive phase elements for laser beam shaping," *Optics Letters*, vol. 27, no. 16, pp. 1463–1465, 2002.